(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 239 061 A2**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **87104288.3**

(51) Int. Cl.⁴: **G06K 9/72**

(22) Date of filing: **24.03.87**

(30) Priority: **24.03.86 JP 63973/86**

(43) Date of publication of application:
30.09.87 Bulletin 87/40

(84) Designated Contracting States:
DE FR GB

(71) Applicant: **Oki Electric Industry Company, Limited**
7-12, Toranomon 1-chome Minato-ku
Tokyo 105(JP)

(72) Inventor: **Nagasawa, Hikoshi c/o Oki Electric Ind. Co. Ltd.**
7-12, Toranom 1-chome Minato-ku
Tokyo(JP)
Inventor: **Ito, Kazuo c/o Oki Electric Ind. Co. Ltd.**
7-12, Toranom 1-chome Minato-ku
Tokyo(JP)
Inventor: **Horii, Shigeru c/o Oki Electric Ind. Co. Ltd.**
7-12, Toranom 1-chome Minato-ku
Tokyo(JP)
Inventor: **Tanoshima, Katsuhide c/o Oki Electric Ind. Co. Ltd**
7-12, Toranom 1-chome Minato-ku
Tokyo(JP)

(74) Representative: **Betten, Jürgen Dipl.-Ing.**
Betten & Gleiss Patentanwälte
Reichenbachstrasse 19
D-8000 München 5(DE)

(54) Optical character reader apparatus and optical character reading method.

(57) An optical·character reader apparatus and reading method is provided to optically read character written on a document card to extract as a word a character group read as such for thereby collating for recognition the extracted word with words previously registered. Such a word is rendered to the collation for each character constituting that word. Thereupon, when the number of constituent characters constituting the extracted word is greater than a prescribed number, it is judged whether or not each constituent character is a vowel, and, if so, removed from an object to be collated, and otherwise rendered to the collation with respect to characters other than vowels. As a result, a word which is found to have the largest number of coincident characters among previously registered words, is selected as a candidate word and delivered as the concerning word read from the document card.

EP 0 239 061 A2

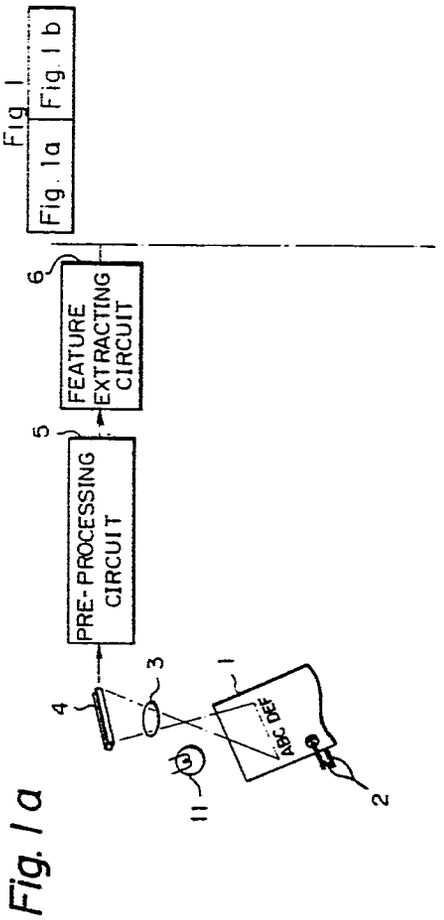


Fig. 1
Fig. 1a
Fig. 1b
PRE-PROCESSING CIRCUIT
FEATURE EXTRACTING CIRCUIT
ABC DEF
1
2
3
4
5
6
11

Fig. 1 a

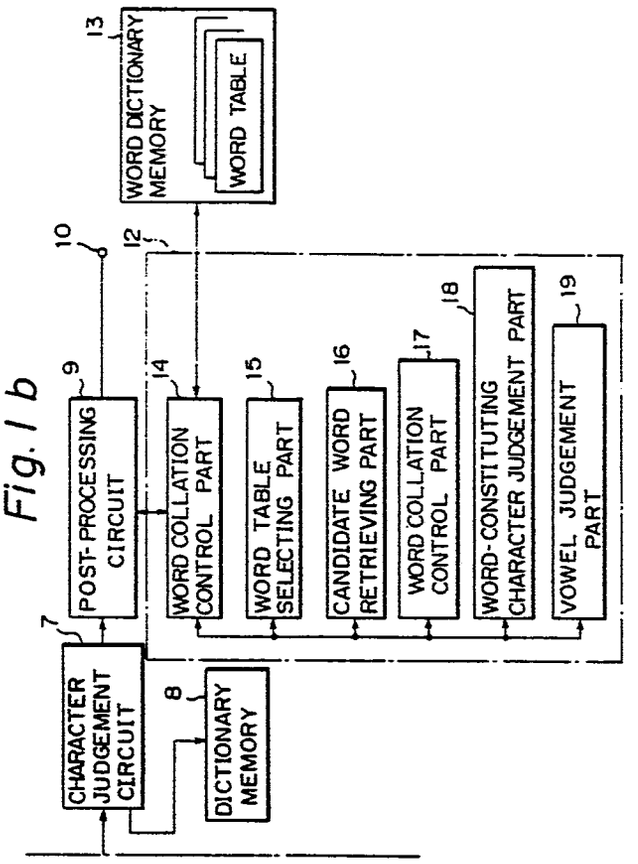

POST-PROCESSING CIRCUIT
CHARACTER JUDGEMENT CIRCUIT
DICTIONARY MEMORY
WORD DICTIONARY MEMORY
WORD TABLE
WORD COLLATION CONTROL PART
WORD TABLE SELECTING PART
CANDIDATE WORD RETRIEVING PART
WORD COLLATION CONTROL PART
WORD-CONSTITUTING CHARACTER JUDGEMENT PART
VOWEL JUDGEMENT PART
7
8
9
10
12
13
14
15
16
17
18
19

Fig. 1 b

OPTICAL CHARACTER READER APPARATUS AND OPTICAL CHARACTER READING METHOD

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to an optical character reader (hereinafter referred to as OCR) apparatus adapted to yield a picture image signal by illuminating characters written on a document card with light from a light source to render reflected light therefrom to photoelectric conversion to provide an electric signal, extracting the feature of respective characters by taking out the original characters from the picture image signal one at a time, and collating the feature of the character extracted as such with a character pattern previously registered in a dictionary memory. More particularly it relates to a word collating system of an OCR adapted to extract a character group as a word, and collate the word so extracted with a word previously registered in the memory for recognizing characters on a document card.

2. Description of the Prior Art:

A prior OCR is adapted, as disclosed in Japanese Patent Application No. 59-125033, to illuminate alphabetic characters written on a document card with a lamp, and focusing reflected light onto a photoelectric converter sensor through a lens to convert the picture image signal on the document card into an electric signal as an output from the sensor.

In succession, a pre-processing circuit takes out a fractional picture image corresponding to one character from the resulting line image and transmits it to a feature extracting circuit. The feature extracting circuit executes the so-called recognition algorithm adapted to extract the features of a character line or a background in conformity with a predetermined procedure. In succession, a character judgement circuit collates the resulting features with those of characters previously stored in a dictionary memory and delivers coincident character codes to a post-processing circuit. Three cases are considered thereupon as the output from the character judgement circuit: a plurality of character codes are yielded; only one character code is yielded; and no character code is yielded. The post-processing circuit will be operated for the above situation as follows: it delivers, when only one character is yielded, it, interpreting the concerning character has satisfactorily been recognized, to the outside; and with no character code being yielded as described above it deliveres a non-recognizable code from its output terminal as no dictionary to recognize the concerning character being existent; and furthermore with a plurality of character codes being yielded this situation may frequently be produced when the concerning character pattern well resembles another character patterns, and hence the concerning character pattern can not be fixed to at least one character code but have a pluraliy of character candidates therefor. There is a method to abandon unnecessary candidates among those candidates for selecting only one character code therefrom, by making use of previously known information indicating a certain character is not written on the document card adjoining to another certain character on the basis of the context thereof. For example, when "U" and "V" is yielded as those character candidates in an Engligh sentense with a character "Z" located in front of them, provided the "V" is regarded as being improper as the character to be located just behind the "Z", it is abandoned with a result of "U" being selected, and thus the "U" code is delivered from an output terminal. Thereupon, information concerning combinations of characters not adjoining to the concerning character code before and behind it can be employed at need by storing them as a table.

In addition, another method is also known to judge whether or not the candidate character is proper by noting adjacent two or three characters to previously provide the frequency of occrurrences of combinations of those characters (this is a known method called 2-gram, 3-gram, and generally n-gram).

Still another method is also known wherein a word is extracted from a character group read by a character judgement circuit to collate it with a word previously stored for judgement at a word level. Namely, character codes of characters constructing read words are collated in succession with those constructing stored words, and a certain stored word having the largest numbers of coincident character codes among those stored words can be judged as the read word.

However, these OCRs to determine the concerning character relying upon such methods as described above suffer from the following problems:

(a) The first method to previously store some information concerning combinations of characters not adjoining the concerning character before and behind it, and to abandon characters corresponding to the above information among a plurality of character candidates can not judge those candidates as

being proper because only those combinations having the context of characters, which can obviously be prohibited, are provided as information, provided a plurality of characters not corresponding to those combinations are presented as the associated candidates.

(b) The second method to store the frequency of occurrences of combined characters for judgement of the concerning character on the basis of the above frequency can not judge characters having similar occurrence frequency as being proper, and furthermore it may judge, instead of a character actually written on a document card but with less occurrence frequency, another character having higher occrurrence frequency as being proper by mistake.

(c) The third method to extract any word for collation thereof with previously stored words takes much time for the collation with the word composed of many characters because it has an enormous amount of words to be collated.

SUMMARY OF THE INVENTION

In view of the drawbacks of the prior arts, it is an object of the present invention to provide an optical character reader (OCR) apparatus and reading method capable of more securely judging, when a character on a document card can not be recognized and hence a plurality of character candidates are yielded, that character on a document card to be proper by uniquely selecting it from those candidates.

Another object of the present invention is to provide an optical character reader apparatus and reading method capable of character recognition even when character recognition can not be effected and hence no character candidate is presented.

Still another object of the present invention is to provide an optical character reader apparatus and reading method capable of reducing processing time of collation by effecting it in a unit of a word in a character group and further effecting the same after removing vowels involved in characters constituting that word.

To achieve the above objects, an optical character reader apparatus (hereinafter, referred to as an OCR apparatus and reading method according to the present invention is adapted to illuminate characters on a document card with light from a light source to render reflected light therefrom to photoelectric conversion for thereby providing electric signal to yield a picture image signal, extracting the feature of each of characters by taking out the characters one at a time from the above picture image signal, and collating the feature so extracted with character patterns previously stored in a dictionary memory, for character recognition, and furthermore extracting a character group between successive two spaces as a word to collate the word extracted as such with words previously stored, for recognition of words on a document card.

Moreover, the OCR apparatus and reading method according to the present invention includes a word table selecting part for selecting a proper word table in a word dictionary memory correspondingly to the number of characters constituting the extracted word; a word-constituting character number judgement part for judging whether or not the number of characters of the extracted word is greater than a prescribed number; a vowel judgement part for judging, when the above numbers of characters constituting the extracted word is judged to be greater than a prescribed number in the word-constituting character number judgement part, whether or not each of those characters constituting the extracted word is a vowel, and removing, if so, a constituting character judged to be a vowel from a collation object; a word collating part for collating constituent characters constituting the extracted word after passing through the vowel judgement part with those characters constituting each word in a word table selected in the word table selecting part for evaluating the number of coincident characters; a candidate word retrieving part for selecting a candidate word having the largest number of coincident characters evaluated in the word collating part; and a post-processing part for delivering to the outside the candidate word selected in the candidate word retrieving part.

The above and other objects, features and advantages of the present invention will become more apparent from the following description when taken in conjunction with the accompanying drawings in which a preferred embodiment of the present invention is shown by way of illustrative example.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram illustrating an embodiment of an OCR apparatus according to the present invention;

Fig. 2 is a flow chart illustrating operation of a word collator circuit 12 of Fig. 1;

Fig. 3 (a) is a view illustrating a format of a memory table storing read results therein;

Fig. 3 (b) is a view illustrating a format of a word table, which previously stores words therein, in a word dictionary memory 13;

Fig. 4 is a block diagram illustrating a microprocessor circuit for executing functions and controls of a post-processing circuit 9 and a word collator circuit 12; and

Fig. 5 is a block diagram illustrating an I/O circuit connected to the microprocessor circuit of Fig. 4.

DESCRIPTION OF THE PREFERRED EMBODIMENT

In what follows, an embodiment of an optical character reader apparatus according to the present invention will be described with reference to the accompanying drawings.

Fig. I illustrates an OCR apparatus according to the present invention.

As shown in the same figure, designated at 1 is a document card on which characters such as those in an English sentence are written mannually or by printing; 2 is a conveyor mechanism for conveying the document card 1 in conformity with a predetermined format; 3 is a focusing lens for focusing reflected light onto a photoelectric converter sensor 4 from an illumination lamp 11 (described later) on the document cart 1 and reflected therefrom; 4 is the photoelectric converter sensor for rendering the reflected light from the document card 1 to photoelectric conversion into an electric signal; 5 is a pre-processing circuit for converting an electric signal from the photoelectric converter sensor 4 to binary values to yield a binary picture image signal and taking out a picture image corresponding to one character therefrom; 6 is a feature extracting circuit for extracting the feature of a character in conformity with a predetermined procedure; 7 is a character judgement circuit described later for collating a dictionary memory 8 having feature data of characters previously registered therein with feature data being an output from the feature extracting circuit 6 and delivering a character code of a coincident character as a result of the collation so effected; 8 is a dictionary memory for previously registering therein character feature data while corresponding the character feature data to character codes of those characters, 9 is a post-processing circuit for effecting editing and output control, etc., corresponding to the judgement result in the character judgement circuit 7; 10 is an output terminal; and 11 is an illumination lamp.

In succession, operation of the OCR apparatus shown in Fig. 1 will be described.

First, with the document card 1 being travelled by means of the converyor mechanism 2, and illuminated with the illumination lamp 11, reflected light from the document card 1 is focused onto the photoelectric converter sensor 4 through the focusing lens 3 to yield as a sensor output a picture image signal on the document card 1 in a form of electric signal. The conveyor mechanism 2 is controlled by a control circuit (not shown), and such a picture image signal corresponding to a prescribed region on the document card 1 is written into a line image buffer memory (not shown) in the pre-processing circuit 5.

Then, the pre-processing circuit 5 takes out a picture image corresponding to one character from the resulting line picture image and transmits it to the feature extracting circuit 6. The feature extracting circuit 6 executes the so-called recognition algorithm to extract the contour of character lines and the feature of a background portion in conformity with a predetermined procedure. The character judgement circuit 7 collates the dictionary memory 8 having a dictionary for character recognition previously registered with the resulting feature, and delivers a coincident character code to the post-processing circuit 9. There are hereupon expected three types of cases for the output contents from the character judgement circuit 7: first one with a plurality of character codes being yielded; second one with only one character code being yielded; and third one with no character code being yielded. If with, thereupon, only one character code being yielded, the concerning character can be said to be recognized satisfactorily, while with no character code being yielded it is suggested that no dictionary to recognize the concerning character is existent in the document memory 8. In addition, with a plurality of character codes being yielded, it means that they are likely to be frequently produced when the concerning character pattern bears close resemblences to another character patterns, so as not to permit at least one character code to be fixedly yielded but to permit a plurality of character candidates to be yielded.

Successively, the post-processing circuit 9 extracts a word from a character group delivered in succession from the character judgement circuit 7 for effecting collation in a unit of this word. Moreover, since one of characters constituting the word has a plurality of character codes as described above, each of the plurality of character codes is processed as corresponding to the above one character.

In the following, operation of such word collation will be described.

Referring to Fig. 1, designated at 12 is a word collator circuit, 13 is a word dictionary memory, 14 is a word collation control part, 15 is a word table selecting part, 16 is a candidate word retrieving part, 17 is a word collator part, 18 is a word-constituting character number judgement part, and 19 is a vowel judgement part. Fig. 2 is a flow chart illustrating operation of the word collator circuit 12. In addition, Fig. 3 (a) is a view illustrating a format of a memory table in which read results are stored,

while Fig. 3 (b) is a view illustrating a format of a word table in the word dictionary memory 13, in which table words such as those in an English sentence are previously stored.

As, hereupon, illustrated in Fig. 3 (a), designated at 20 is a memory table for storing therein read results yielded by reading a character group written as a single word in English sentence, etc., by means of the OCR apparatus in conformity with the above-described procedure, and 21-1, 21-2, ..., 21-i, ..., 21-N are areas for storing read results of respective characters constituting a word with C0, C1, ..., Ci, ..., CN respectively indicating the contents of the storages thereof. Hereupon, N indicates the number of characters of a word. Moreover, i is a pointer indicative of the character number in the table 20. Furthermore, as illustrated in Fig. 3 (b), designated at 30 is a table, and 31-1, ..., 31-k, ..., 31-M are respectively storage locations of words registered in the table 30. The table 30 is therefore adapted to register M words therein.

Noting now the kth word 31-k, this word consists of N character, each character being designated at $W0^k$, $W1^k$, ..., Wi^k, ..., WN^k. Wi^k here represents the ith constituent character of the kth word registered in the word table.

Here, although Fig. 3 illustrates only the word table constituted by N characters, another many word tables may instead be expected depending on the numbers of constituent characters because of various types of constituent characters constituting words being considered, but they are here not shown in the figure for brevity. Accordingly, the present OCR apparatus is adapted to select a proper word table responsibly to the number of read characters and in a corresponding manner thereto.

First, the post-processing circuit 9 extracts as a word a character group disposed between adjacent spaces, and delivers it to the word collation control part 14 in the word collating circuit 12. Assuming here this word to be constructed with N characters, the word table selecting part 15 selects a word table composed of N characters in the word dictionary memory 13 (step 100). The word collator part 17 initializes a coincidence number counter "COIN" counted up when the character pointer i, word pointer k, ith character constituting a word in each word, and ith character as a read result are coincident with each other (step 101 and 102). In succession, the candidate word retrieving part 16 takes out the ith character Ci as a read result (step 103). Successively, the word-constituting character number judgement part 18 compares the number N of characters constituting a word as a read result with a constant D2 to thereby judge whether or not the former number N is greater than the latter constant D2 (step 104), and the control advances, if $N < D2$, to step 106. In addition, if $N \geq D2$, the vowel judgement part 19 judges whether or not the ith character Ci as the read result is a vowel (step 105), and if not, the control advances to step 106. While, if so, the judgement part 19 regards the ith character Ci to be a candidate character and the control advances to step 114 described later after removing it from a collation object.

In succession, step 106 judges the attribute of Ci. Here, the attribute means the number of character codes yielded as a read result, which number is indicated by S. This value S is, although not shown in the figure, known upon reading and hence is assumed to be stored. Thereafter, those candidate words yielded as described above and read results are rendered to the concerning judgement. Provided as a result of step 106 the character code number is $s=0$ (this means no candidate word is yielded), the control advances to step 114 described later. And, provided the number is $s=1$ (this means only one candidate word is yielded), it is checked that the read result Ci and the ith constituent character Wi^k of the kth word in the word table are coincident with each other (step 107), and if coincident the coincident counter "COIN" is counted up (step 108), and otherwise the control advances to the step 114. Moreover, if $s \neq 0$, and $s \neq 1$, which means a plurality of character codes are yielded as a read result, the candidate word retrieving part 16 searches a plurality of numbers of candidates from $\ell = 0$ to s (steps 109, 110, 111, and 112). If there is any coincident code thereamong, the coincident number counter "COIN" is counted up (step 113), and otherwise the control advances to step 114. In the step 114, the character number counter i is counted up, and if this counter i is less than N characters, the control returns to the step 103 to judge the next character with the same operation (step 115). Hereby, all characters involved in the concerning word are rendered to the judgement.

In succession, the contents of the coincident number counter "COIN" are examined and compared with the constant D1 (step 116), and if the former is less than the latter, the control advances to step 118, and otherwise, the contents of the "COIN" and k are reserved (step 117). Those operations in the steps 116 and 117 are based on a principle, say, less coincident words are abandoned at that time. Here, the reserved value of k is the number of the coincident words, and means this word is registered as one of candidate words clearing the judgements up to now. In addition, a plurality of areas are available for the above reservation, which are adapted to register thereon all words having coincident numbers more than the constant D1, which value may be 10 at most. Moreover, D1, although assumed to be a constant,

may be made variable dynamically as a function of the word-constituting character number N every time at need. And, furthermore, it is judged whether or not the operations effected until now are executed for all words involved in the concerning word table (steps 118, 119) and if not the control returns to the step 102, and otherwise advances to step 120.

In succession, in the steps 120 and 121, a collation check in a word unit between the concerning word table and the read result is finished, and number of words, in the concerning word table having the coincident number more than D1 as well as the number of coincidences are preserved. Accordingly, the number of a word having the maximum coincident number is selected among the reserved candidate word group (step 120), and a word name corresponding thereto is transmitted to the outside via the post-processing circuit 9.

Hereupon, when the number of a word exerting the maximum number of coincidence is not uniquely determined, the word collation is unsatisfactory, and the result may be processed with some methods: the read result is delivered as it is; a plurality of words having greater numbers of coincidence are delivered; or both of them are delivered. In any case, the selection may be left on an operator. Otherwise, collation for all constituent characters of a word read again including vowels may be conducted.

Hereupon, although the coincident number counter "COIN" was counted up commonly for words, it is effective to count up the counter by weighting it with values to be added thereto depending on the attribute s. For, example, +3 is employed to weight the character code = 1 therewith yielded in step 106, or +1 is employed to weight if plurality of character codes are yielded in step 106. In addition, if no character code is yielded in the step 106, -1 may be applied as a penalty. This is one method to judge these words in terms of the total of scores for adopting in preference only one read result being yielded.

In the following, exemplary circuits of the post-processing circuit 9, word collating circuit 12, and word dictionary memory 13 to concretely execute the operation described above will be explained with reference to Figs. 4 and 5.

First, referring to Fig. 4, illustrating a microprocessor circuit to execute functions and controls of the post-processing circut 9 and the word collating circuit 12, designated at 201 is an arithmetic logic unit (hereinafter referred to as ALU) to which Advanced Micro Devices Inco., made AM 29116 and the like is fitted, 202 is a microprogram memory, and a microprogram stored therein is employed to operate the microprocessor circuit and an I/O circuit described later, 203 is an I bus composed of sixteen signal lines for interconnecting the above ALU 201 and the microprogram memory 202. The ALU 201 executes arithmetic operation in conformity with an instruction inputted from the IBUS 203. Likewise, designated at 204 is a sequencer providing addresses of the microprogram memory 202, which consists of four elements of AM 2911 for example. In addition, the ALU 201 is connected with the I/O circuit described later through a YBUS 205 for data delivery. Designated at 206 is a decoder for generating SEL1 signal 207 to SEL 4 signal 210 from two signals on the YBUS 205.

Here, a numerical value in ( ) indicates the number of signal lines.

Fig. 5 illustrates the I/O circuit connected to the Y bus 205 of the microprocessor circuit.

Designated at 211 is a flip-flop for storing therein a start signal from the character judgement circuit 7, the start signal being delivered to the Y bus 205 with use of the SEL 1 signal 207, and 212 is a register for storing a candidate character code issued from the character judgement circuit 7, data in this resister 212 being delivered to the YBUS 205 with use of the SEL2 signal 208. Moreover, in the word dictionary memory 13, designated at 213 is a memory for storing the concerning words, 214 is an address register for assigning addresses of the memory 213, and 215 is a data register for storing therein data from the memory 213. The address register 214 incorporates any data from the Y bus 205 with use of the SEL3 signal 209, while the data register 215 delivers any data to the Y bus 205 with use of the SEL 4 signal 210.

Although a certain preferred embodiment has been shown and described, it should be understood that many changes and modifications may be made threin without departing from the scope of the appended claims.

Claims

(1) An optical character reader apparatus comprising:

(a) an optical system for converting characters written on a document card to an optical signal and further rendering said optical signal so converted to an electric signal;

(b) a pre-processing circuit connected to said optical system for converting said electric signal into a binary picture image signal, and taking out a picture image corresponding to one character from said binary picture image signal;

(c) a feature extracting circuit connected to said pre-processing circuit for extracting the features of the character in conformity with a predetermined procedure;

(d) a character judgement circuit connected to said feature extracting circuit for effecting the collation of the concerning character;

(e) a dictionary memory connected to said character judgement circuit for previously registering feature data of characters while corresponding said data to character codes of said characters, and further permitting said character judgement circuit to collate said feature data from said feature extracting circuit with said feature data registered in said dictionary memory and to deliver character codes of coincident characters as a result of said collation;

(f) a post-processing circuit connected to said character judgement circuit and an output terminal for effecting editing and output control, etc., corresponding to said result of judgement in said character judgement circuit; and

(g) a word collator circuit connected to said post-processing circuit for effecting word collation in cooperation with said circuit.

(2) An optical character reader apparatus according to claim 1, wherein said optical system includes:

(a) a document card conveyor mechanism for conveying a document card in conformity with a predetermined format;

(b) an illumination lamp for illuminating the document card;

(c) a focusing lens for focusing reflected light from the document card; and

(d) a photoelectric conversion sensor for rendering the reflected light from the document card to photoelectric conversion to provide an electric signal.

(3) An optical character reader apparatus according to claim 1, wherein said word collator circuit includes:

(a) a word dictionary memory for previously registering words in word tables involved therein while classifying said words for every number of characters constituting said word;

(b) a word collation control part connected to said post-processing circuit and said word dictionary memory for controlling the word collation;

(c) a word table selecting part connected to said word collation control part for selecting any of said word tables involved in said word dictionary memory responsibly to the number of characters constituting an extracted word;

(d) a word-constituting character judgement part connected to said word collation control part for judging whether or not the number of characters constituting the extracted word is greater than a prescribed number;

(e) a vowel judgement part connected to said word collation control part for judging whether or not respective constituent characters of the extracted words are vowels when said number of characters constituting the word extracted in said word constituting character number judgement part is greater than a prescribed number, and removing, if so, a constituent character judged to be a vowel among said constituent characters from the object to be collated;

(f) a word collator part connected to said word collation control part for collating the constituent characters of the extracted word yielded through said vowel judgement part with constituent characters of each word in said word table selected in said word table selecting part and for thereby evaluating the number of coincident characters; and

(g) a candidate word retrieving part connected to said word collation control part for selecting the candidate word having the largest number of the coincident characters in said word collator part:

(4) An optical character reader apparatus according to claim 1, wherein the functions and controls of said post-processing circuit and said word collator circuit are executed by a microprocessor circuit.

(5) An optical character reader apparatus according to claim 3, wherein the functions and controls of said post-processing circuit and said word collator circuit are executed by a microprocessor circuit.

(6) An optical character reader according to claims 1, 3, 4, 5 wherein said microprocessor includes: an arithmetic logic unit for effecting arithmetic operation; a microprogram memory for operating said microprocessor circuit in conformity with a microprogram stored therein; an I bus composed of 16 signal lines for mutual connection of said arithmetic logic unit and said microprogram memory; a sequencer connected to said microprogram memory for providing addresses of said microprogram memory; a Y bus connected to said arithmetic logic unit for use in data delivery; an I/O circuit connected to said Y bus; and a decoder connected to said Y bus and said sequencer for generating a selection signal from two signals on the Y bus.

(7) An optical character reader apparatus according to claim 6, wherein said I/O circuit includes: a flip-flop connected to said character judgement circuit for storing a start signal from said character judgement circuit and transmitting it to said Y bus by the SEL 1 signal; a candidate character code register connected to said character judgment circuit and the Y bus for storing a candidate character code delivered from said character judgement circuit and transmitting it to the Y bus by the SEL 2 signal; and, in said word dictionary memory, a memory for storing words, an address

register for assigning any address of the memory, and a data register for storing data from said memory, said address register incorporating therein data from said Y bus based upon the SEL 3 signal while said data register delivering the concerning data to the Y bus based upon the SEL 4 signal.

(8) An optical character reading method comprising the steps:

(a) registering feature data of characters in a dictionary memory while corresponding said data to character codes of said characters;

(b) converting characters written on a document card to an optical signal and further rendering said optical signal so converted to an electric signal;

(c) converting said electric signal into a binary picture image signal, and taking out a picture image corresponding to one character from said binary picture image signal;

(d) extracting the features of the character in conformity with a predetermined procedure;

(e) effecting the collation of the concerning character with said feature data registered in said dictionary memory and delivering character codes of coincident characters as a result of said collation;

(f) effecting editing and output control, etc., corresponding to said result of collation; and

(g) in a word collator circuit effecting word collation with the result of said character collation.

(9) An optical character reading method according to claim 1, wherein said effecting the word collation comprises the following steps:

(a) previously registering words in word tables involved therein in a word dictionary memory while classifying said words for every number of characters constituting said word;

(b) controlling the word collation by a word collation control part;

(c) selecting any of said word tables involved in said word dictionary memory responsibly to the number of characters constituting an extracted word;

(d) judging whether or not the number of characters constituting the extracted word is greater than a prescribed number;

(e) judging whether or not respective constituent characters of the extracted words are vowels when said number of characters constituting the word is greater than a prescribed number, and removing, if so, a constituent character judged to be a vowel among said constituent characters from the object to be collated;

(f) collating the constituent characters of the extracted word yielded through said vowel judgement with constituent characters of each word in said selected word table and for thereby evaluating the number of coincident characters; and

(g) selecting the candidate word having the largest number of the coincident characters in said word collating.

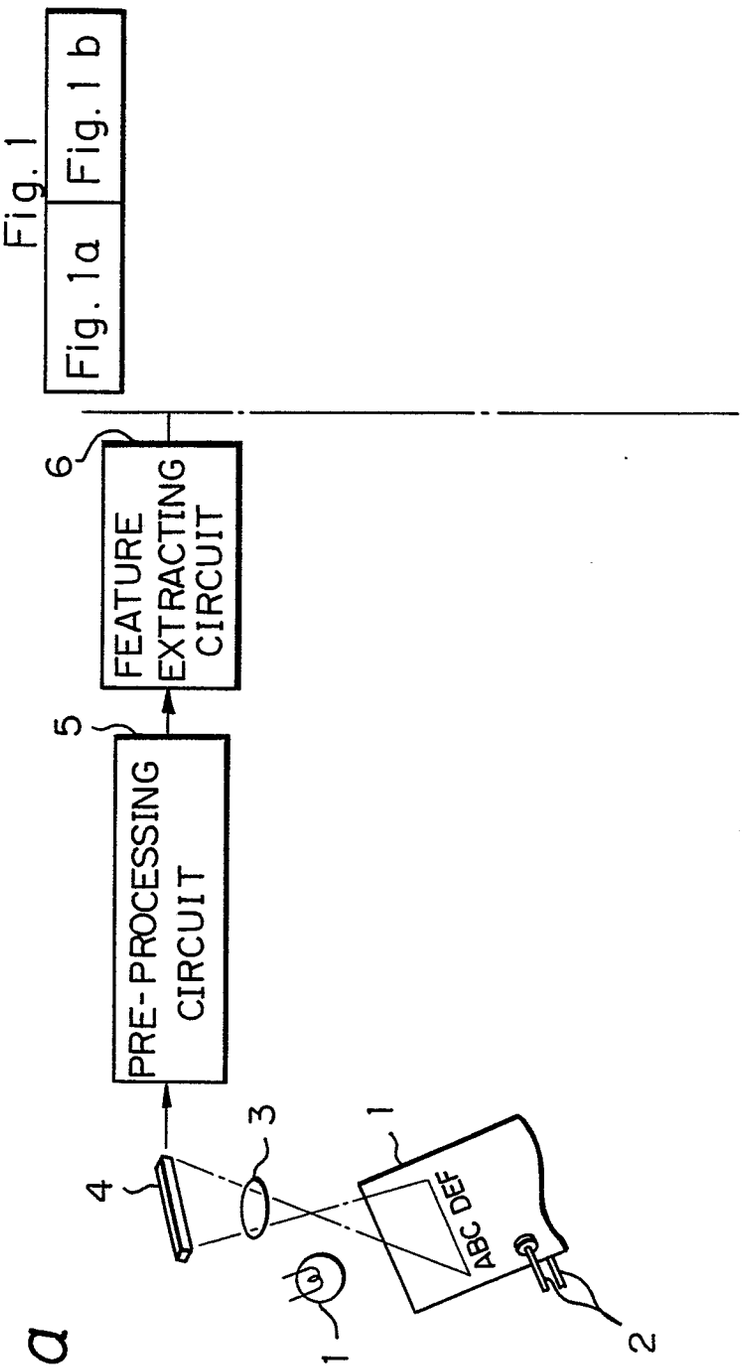


Fig. 1
Fig. 1a
Fig. 1 b
6
FEATURE EXTRACTING CIRCUIT
5
PRE-PROCESSING CIRCUIT
4
3
1
ABC DEF
11
2

Fig. 1a

Fig. 1 b.

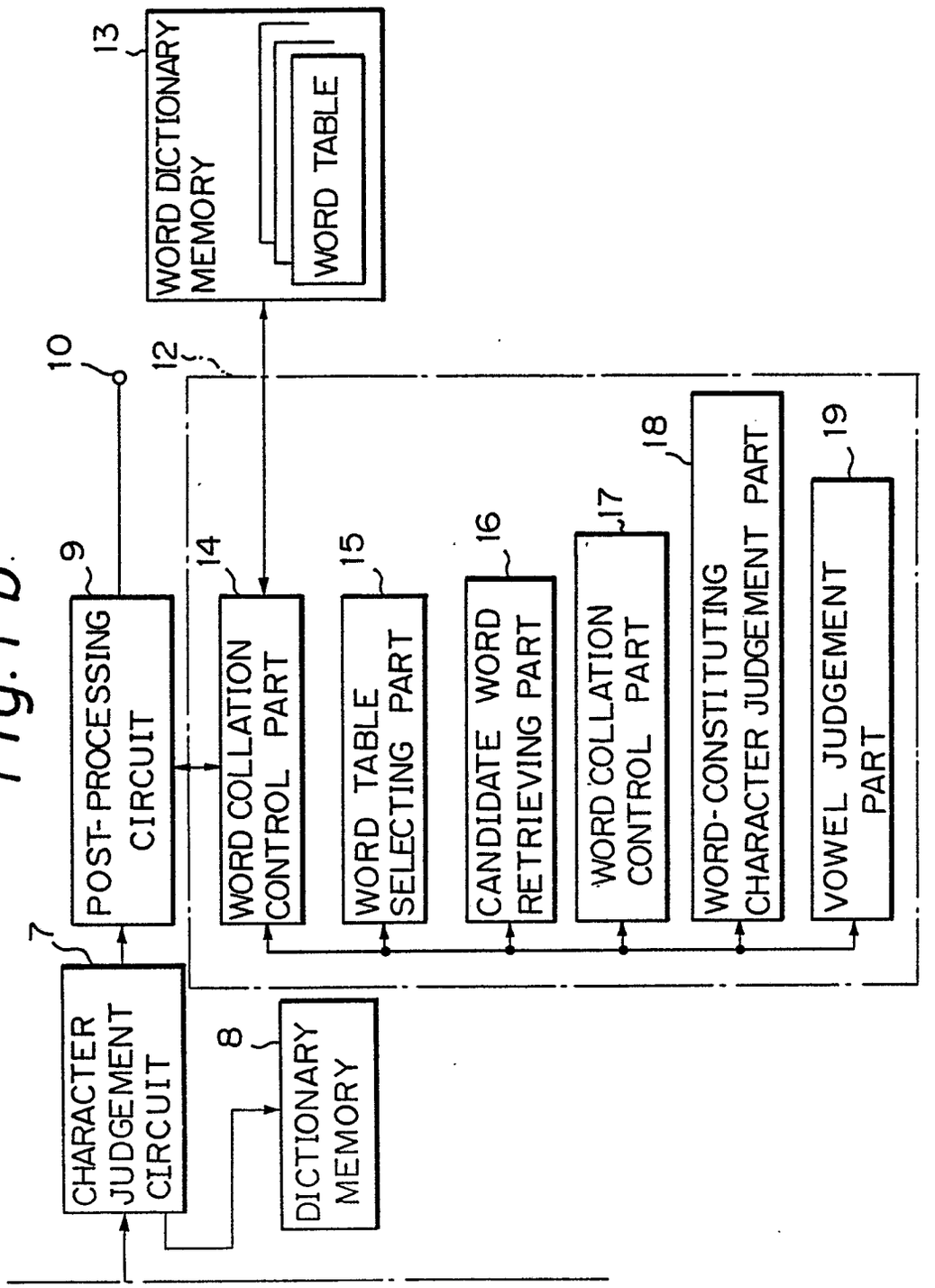

CHARACTER JUDGEMENT CIRCUIT
7
DICTIONARY MEMORY
8
POST-PROCESSING CIRCUIT
9
10
WORD DICTIONARY MEMORY
13
WORD TABLE
WORD COLLATION CONTROL PART
14
WORD TABLE SELECTING PART
15
CANDIDATE WORD RETRIEVING PART
16
WORD COLLATION CONTROL PART
17
WORD-CONSTITUTING CHARACTER JUDGEMENT PART
18
VOWEL JUDGEMENT PART
19

Fig. 2 a

Fig. 2

Fig. 2 a

Fig. 2 b

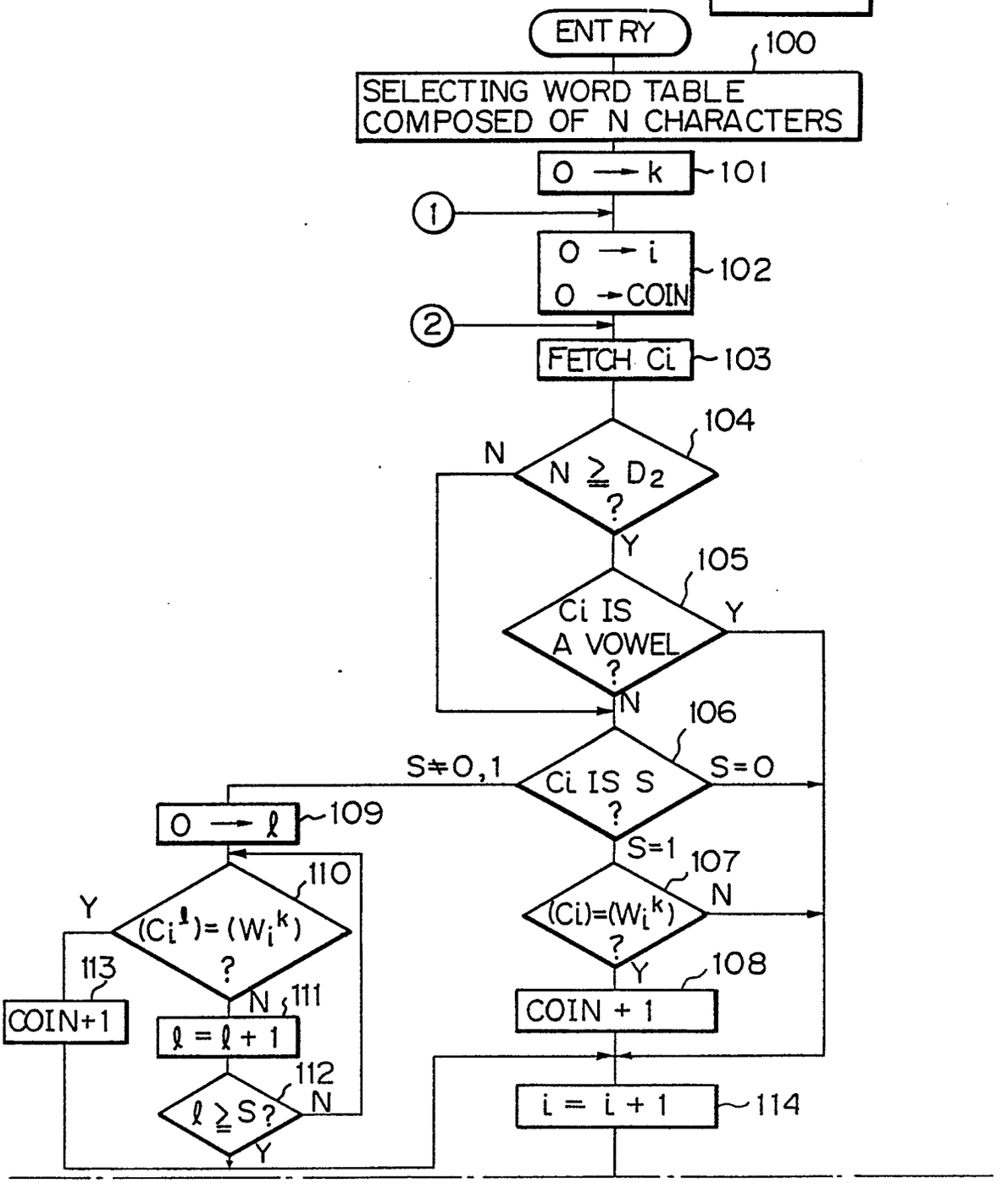


Fig. 2 b

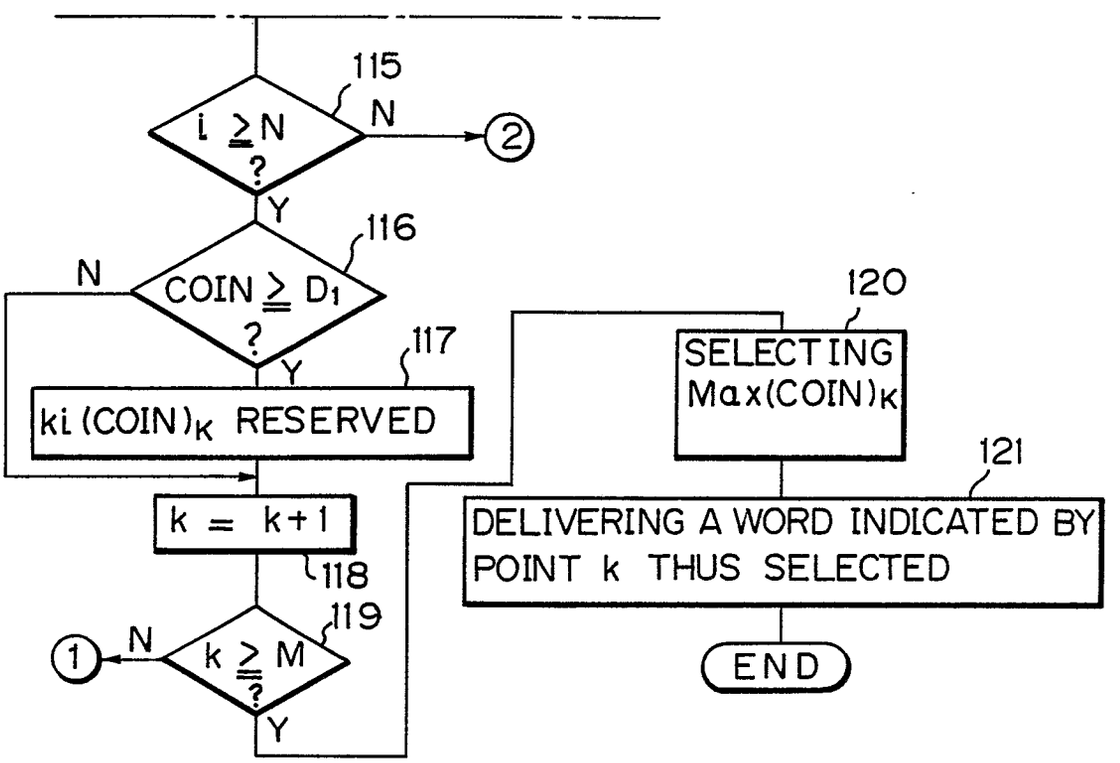
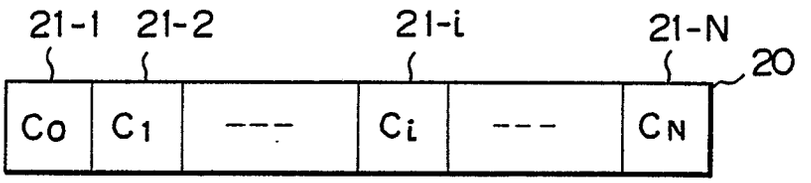

115
i ≧ N ?
N
2
Y
116
N
COIN ≧ D1 ?
Y
117
ki (COIN)K RESERVED
k = k+1
118
119
1
N
k ≧ M ?
Y
120
SELECTING Max(COIN)K
121
DELIVERING A WORD INDICATED BY POINT k THUS SELECTED
END

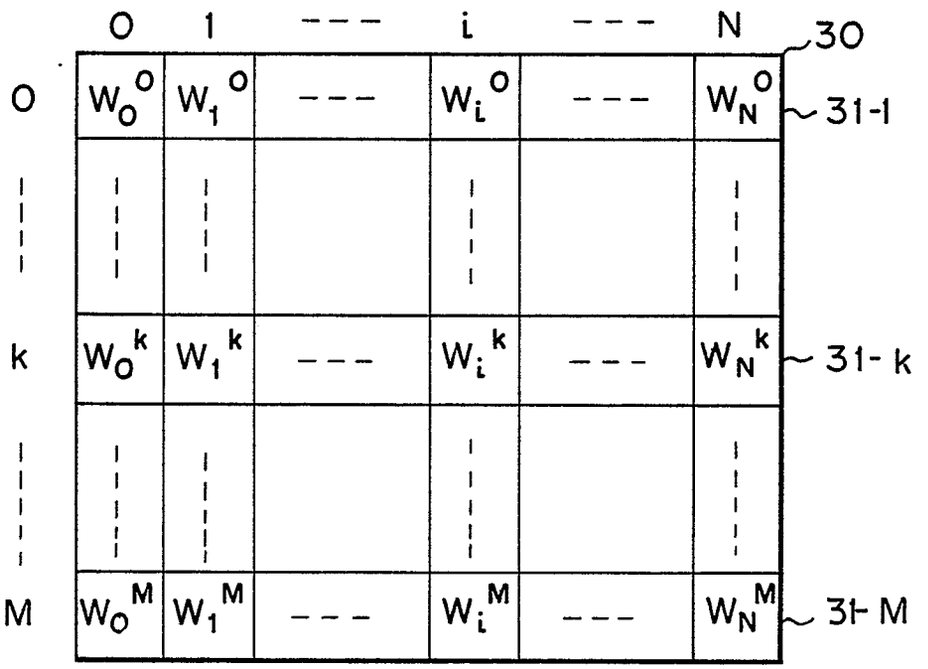
Fig. 3 a


21-1
21-2
21-i
21-N
20
Co
C1

Ci

CN

Fig. 3 b


0
1
---
i

N
30
0
k
M
W0 0
W1 0
Wi 0
WN 0
31-1
W0 k
W1 k
Wi k
WN k
31- k
W0 M
W1 M
Wi M
WN M
31- M

Fig. 4

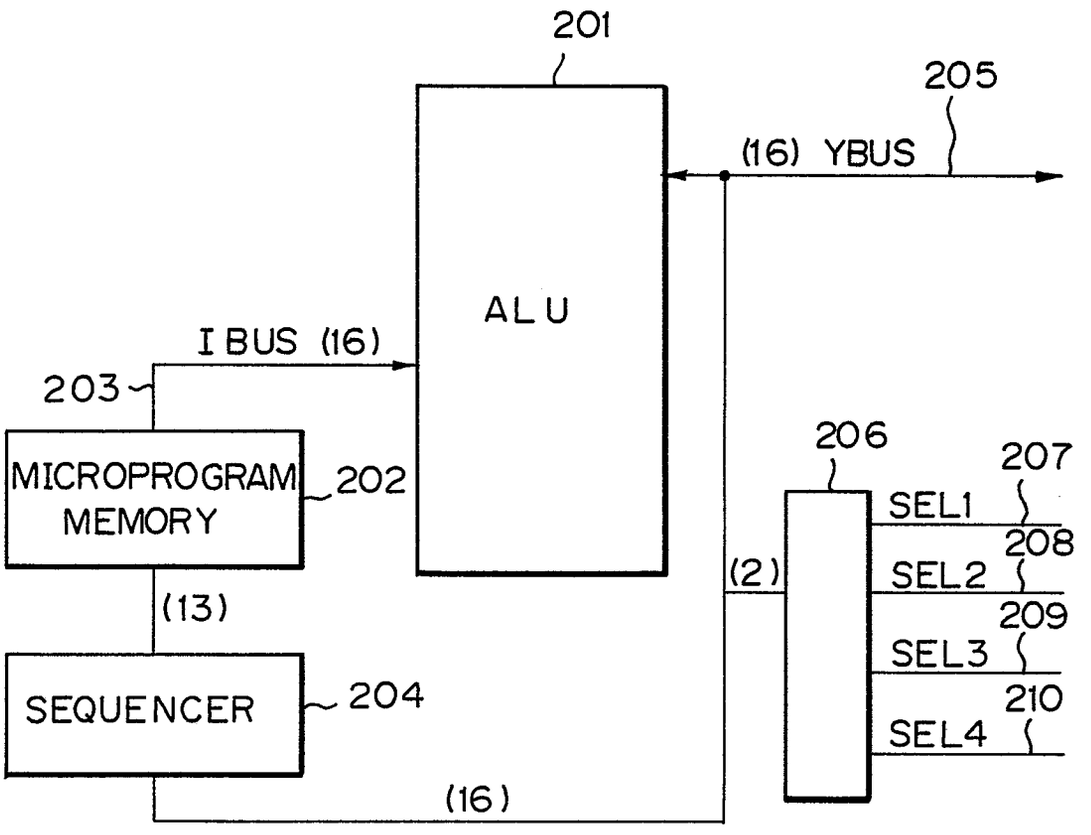
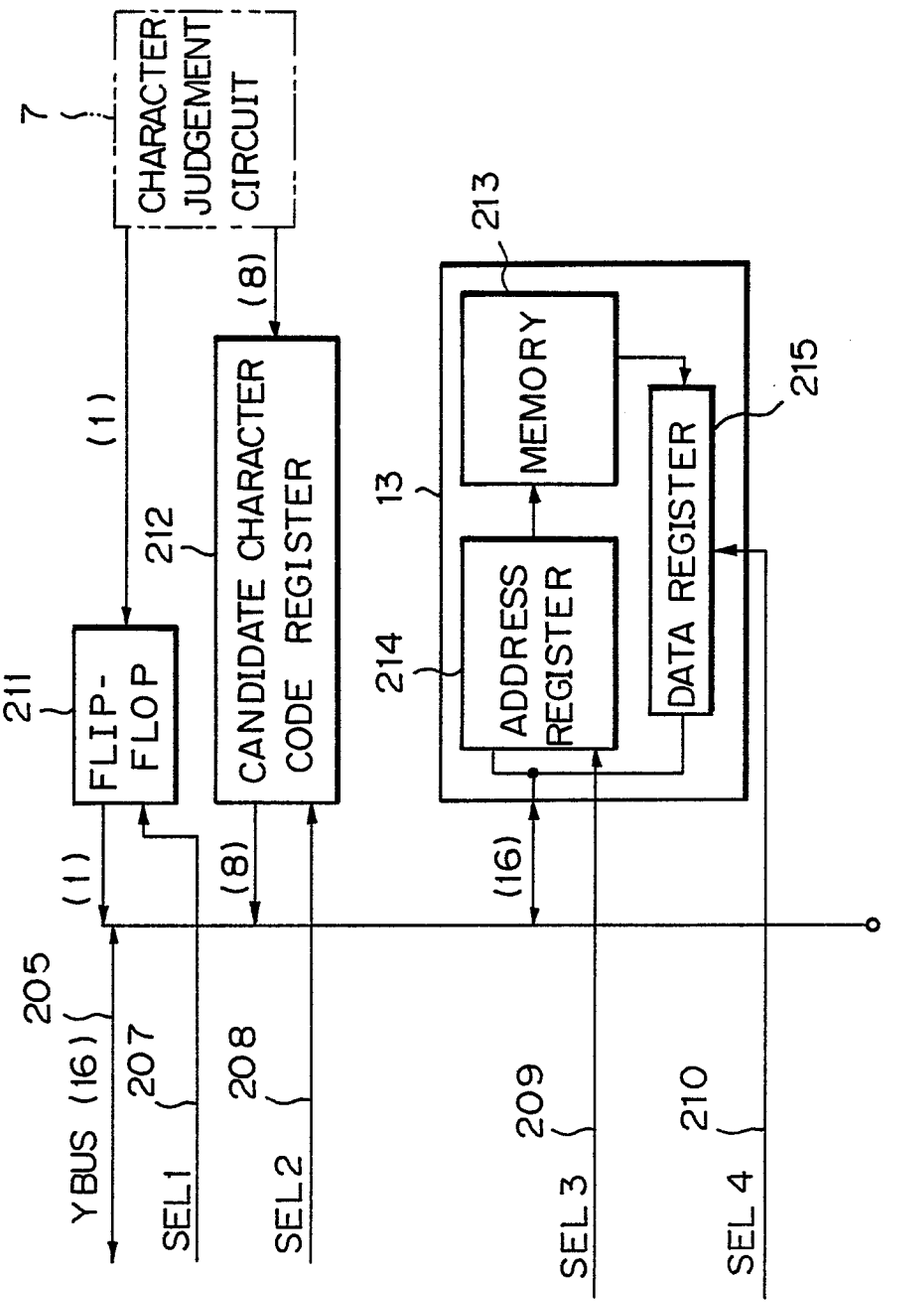

201
205
(16) YBUS
ALU
IBUS (16)
203
MICROPROGRAM MEMORY
202
206
207
SEL1
208
SEL2
(2)
209
SEL3
210
SEL4
(13)
SEQUENCER
204
(16)

Fig. 5


7
CHARACTER JUDGEMENT CIRCUIT
(1)
(8)
211
FLIP-FLOP
212
CANDIDATE CHARACTER CODE REGISTER
213
MEMORY
13
214
ADDRESS REGISTER
215
DATA REGISTER
(1)
(8)
(16)
205
YBUS (16)
207
SEL1
208
SEL2
209
SEL 3
210
SEL 4